Determining a Confidence Factor for Automatic Target Recognition Based on Image Sequence Quality

Gregory J. Power[a]

Air Force Research Laboratory, AFRL/SNAT/Target Recognition Branch,

2010 5th Street, Wright-Patterson AFB, Ohio 45433-7001

Mohammad A. Karim[b]

University of Dayton, Department of Electrical and Computer Engineering,

300 College Park, Dayton, Ohio 45469-0226

ABSTRACT

For the Automatic Target Recognition (ATR) algorithm, the quality of the input image sequence can be a major determining factor as to the ATR algorithm's ability to recognize an object. Based on quality, an image can be easy to recognize, barely recognizable or even mangled beyond recognition. If a determination of the image quality can be made prior to entering the ATR algorithm, then a confidence factor can be applied to the probability of recognition. This confidence factor can be used to rate sensors; to improve quality through selectively preprocessing image sequences prior to applying ATR; or to limit the problem space by determining which image sequences need not be processed by the ATR algorithm. It could even determine when human intervention is needed. To get a flavor for the scope of the image quality problem, this paper reviews analog and digital forms of image degradation. It looks at traditional quality metric approaches such as peak signal-to-noise ratio (PSNR). It examines a newer metric based on human vision data, a metric introduced by the Institute for Telecommunication Sciences (ITS). These objective quality metrics can be used as confidence factors primarily in ATR systems that use image sequences degraded due to transmission systems. However, to determine the quality metric, a transmission system needs the original input image sequence and the degraded output image sequence. This paper suggests a more general approach to determining quality using analysis of spatial and temporal vectors where the original input sequence is not explicitly given. This novel approach would be useful where there is no transmission system but where the ATR system is part of the sensor, on-board a mobile platform. The results of this work are demonstrated on a few standard image sequences.

Keywords: Target Recognition, Image Quality, Probability of Recognition, Image Sequence, Image Degradation

1. INTRODUCTION

For a system that generates electro-optical image sequences, one model of recognition has been defined as a multi-step process involving detection, segmentation, feature extraction, indexing, and matching.[1] Image quality can also be added as an early step toward recognition. After all, based on quality, an image can be easy to recognize, barely recognizable or even mangled beyond recognition. A simple statistical system-level analysis of this multi-step model is shown in figure 1, where the probability of recognition depends on the probability of success at each step. This is stated mathematically as

$$P(Recognition) = P(Match|Index, Feature, Segment, Detect, Quality).$$

This thought process suggests that if detection fails, (i.e. $P(\text{detection}) = 0$) then so does every other step in the multi-step recognition process. However, if detection succeeds, there is still the chance that every other step will fail if segmentation fails (i.e. $P(\text{segmentation}) = 0$). Likewise, if any step fails so will the recognition. However, this statistical thought process also leaves room for degrees of success at each step (i.e. $P(\text{step}) > 0$) which suggests that the probability of recognition can be built based on confidence in the individual steps.

Other author information: (Send correspondence to G.J.Power)
G.J.Power: E-mail: powergj@aa.wpafb.af.mil
M.A.Karim: E-mail: mkarim@engr.udayton.edu

If the input image sequence quality is poor then so may be the detection and/or segmentation and/or feature extraction and so on. Suggesting that quality is a key initial step in the recognition, multi-step process. Thus, quality is added in figure 1 as the first step toward recognition.

A logical goal would be to keep image quality high enough so it does not interfere with the other steps in the recognition process. After all, with articulation of targets and the dynamics of the real world environment, the other recognition steps have enough to worry about. For the most part, the quality aspects of the recognition system have been left to subsystem designers who define quality factors for subsystems such as the optical or transmission subsystem. However, a subsystem quality factor is not sufficient to define the overall impact of quality on the probability of recognition. The actual quality varies based on a variety of factors which may include sporadic system failures, atmospherics, environmental, other analog artifacts, and digital artifacts. This paper suggests that system quality can be transformed into a factor that determines confidence in the capability of the recognition process. Further, unlike other quality metrics, this paper introduces a technique that needs no knowledge of the input imagery to determine the quality of the output image frame.

2. SCOPE OF IMAGE QUALITY PROBLEM

The cause for the perceived poor quality of imagery can be a combination of problems starting from the object being imaged and ending with the eyesight and mind of the person observing. For our purposes, we will assume a typical object (one that is not an ambiguous illusion) and a normal observer. In that case, the cause for poor perceived quality can be anything that occupies the space between the observer and the object. Starting from the object, quality is impacted by the atmosphere and obstructions between the object and the sensor, followed by the optical system of the sensor, the sensor electronics, digitization, compression, software, the transmission system along with atmospherics, the receiving system, the display system and the space between the display and observer (Figure 2). To get a flavor for the scope of the image quality problem, this paper briefly reviews analog and digital forms of image degradation.

2.1. Analog Forms of Image Degradation

Many reasons exist for poor quality within the various subsystems involved in imaging. In the atmosphere, phenomena such as smoke, fog, rain, lightening, sun or even other objects can cause poor quality. In the optical subsystem of either an infrared or a standard TV camera, quality can be impacted by focus, field of view, magnification, optical efficiency, geometrical distortion, optical aberrations and diffraction effects.[2] For the person viewing a CRT screen, the quality of the imagery can be poor due to misaligned controls for contrast, color, horizontal hold and so forth. Analog TV viewers may see noise with periodic components such as herringbone, moirés, and flutter[3] The quality can be poor due to the analog transmission with problems such as a weak signal fading in and out, ghosting, static, or interference. Ghosting is caused by multipath and out-of-phase reception. Fading can be the result of a poor receiver or atmospherics. Static might result because the transmitter is too far away. Interference might be caused by a nearby strong transmitter interfering near the frequency of the TV channel.

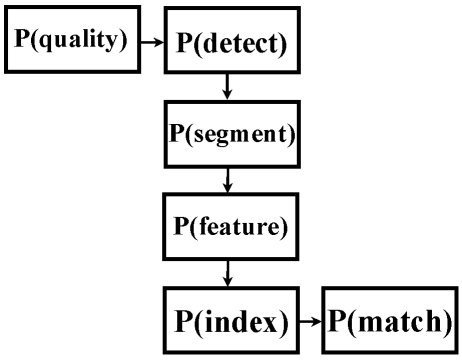

Figure 1. A system analysis shows that the probability of recognition depends on the probability of success in each step of the recognition model.

2.2. Digital Forms of Image Degradation

Some of the poor quality problems of the analog systems are being improved with digital systems. Digital TV images do not exhibit all the same problems due to rejection of amplitude modulation.[3] Also, ghosting and fading in and out is not a typical problem with a digital transmission system. Displays and cameras have been improved with digital circuitry. However, quality degradation for digital systems occur due to the CCD resolution in the digital camera, the display pixel resolution, quantization noise, interference resulting in bit loss and digital transmission coding schemes. The digital transmission systems impact image quality with artifacts such as blocking/tiling, jerkiness, edge busyness and image persistence.[4] This suggests that ever-changing technology offers ever-changing ways not only to improve image quality, but also to degrade image quality.

3. OBJECTIVE QUALITY METRICS FOR TRANSMISSION SYSTEMS

Determining an appropriate objective quality metric is an ongoing research problem. Much research has focused on the transmission systems where bench-top tests can use knowledge of the high quality input imagery to the system and knowledge of the degraded output imagery to produce an objective metric. For Automatic Target Recognition (ATR) systems that use transmission systems, existing objective quality metrics can be used as confidence factors. Metrics that use knowledge of the input and output include metrics based on the mean square error (MSE), metrics based on correlation with subjective assessments, and metrics based on human visual system (HVS) models. This section reviews two common metrics, a common MSE approach known as PSNR and the $\hat{s}$ metric based on correlation with subjective assessment.

3.1. PSNR Metric

A traditional approach for determining an objective quality metric for imagery is the Peak Signal-to-Noise ratio (PSNR). To calculate PSNR, for an image $g(x, y)$ that is somewhat degraded compared to an original image $f(x, y)$, a comparison of the two images can be made by looking at the error,[5] $e(x, y)$ such that

$$e(x, y) = g(x, y) - f(x, y)$$

and the mean square error MSE over the whole image of size M by N is given by

$$MSE = \frac{\sum_{x=1}^{M} \sum_{y=1}^{N} e^2(x, y)}{MN}$$

and, consequently, the PSNR is calculated as

$$PSNR = 20 \log_{10} \frac{[\text{peak value of } g(x, y)]}{\sqrt{MSE}}$$

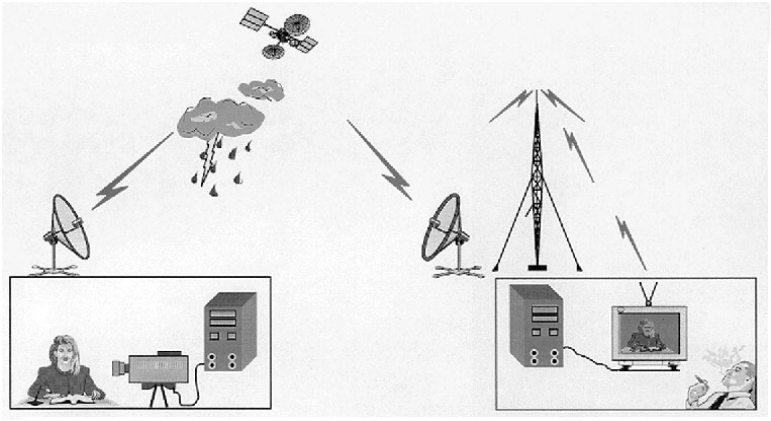

Figure 2. Poor perceived quality can be the result of something that occurs between the observer and the object

where the peak value of $g(x, y)$ is 255 for an 8 bit image.

PSNR has been shown not to be an accurate reflection of the quality perceived by the human visual system (HVS).[6] An example of why PSNR is not an accurate reflection of the HVS is shown in the infrared image of figure 3. One image is degraded with additive zero mean normally distributed noise. The other image is also degraded with the same additive noise after each column of the noise matrix has been ranked in descending order. The result is degraded images with the same PSNR of 11.4 DB. But the HVS suggests different values of quality. In one image the degradation error is spread smoothly throughout the image while in the other image, the degradation error is more random giving a peppered look and the perception of a lower quality image.

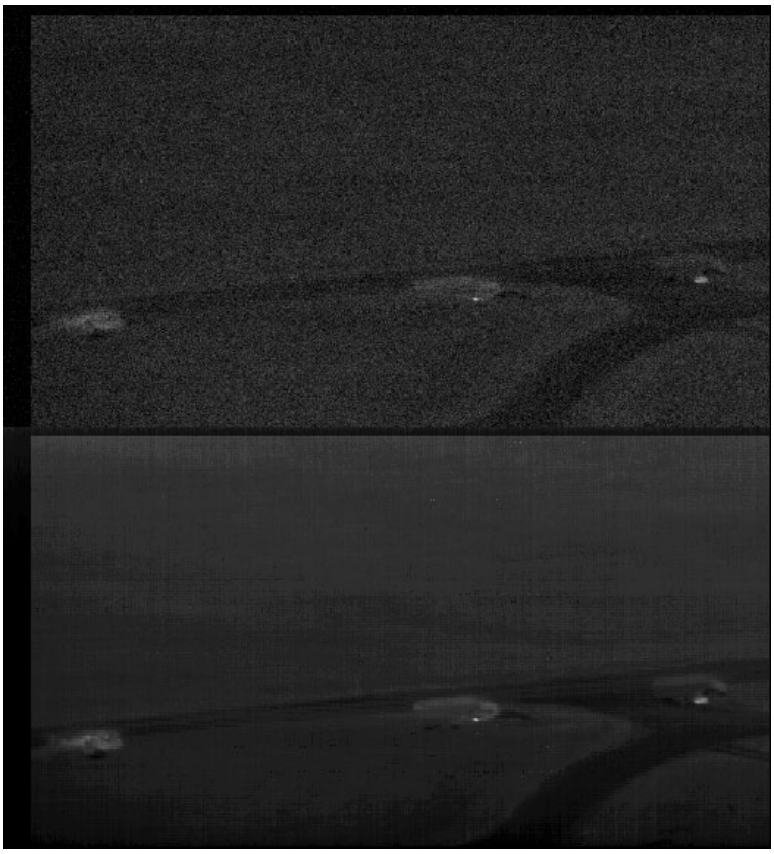

Figure 3. Both images are degraded with additive noise. The human visual system assigns different values of quality to the images although both have the same PSNR of 11.4 DB.

3.2. ITS Metric

Another objective quality metric denoted as $\hat{s}$ was introduced by the Institute for Telecommunication Sciences (ITS).[7] The metric is defined as

$$\hat{s} = 4.77 - 0.922m_1 - 0.272m_2 - 0.356m_3.$$

To define the variables m_1, m_2,and m_3, some symbols need to be defined. The original sequence is defined as O_n and the lower quality sequence is defined as D_n. $SI[O_n]$ is the spatial information content based on the standard deviation of the edge information in the image, O_n. $TI[O_n]$ is the temporal information content based on the standard deviation calculated over the frame difference between successive images. Defining other symbols, STD is standard deviation, STD_{time} is standard deviation over time, MAX is maximum, MAX_{time} is maximum over time, RMS_{time} is root mean square over time, and $CONV$ is the convolution function. Now, the variables in $\hat{s}$ are defined as

$$m_1 = RMS_{time}(5.81|\frac{SI[O_n] - SI[D_n]}{SI[O_n]}|),$$

$$m_2 = f_{time}(0.108MAX\{(TI[O_n] - TI[D_n]), 0\}), \text{ and}$$

$$m_3 = MAX_{time}\{4.23LOG_{10}(\frac{TI[D_n]}{TI[O_n]}), 0\},$$

where

$$f_{time} = STD_{time}\{CONV(x_i, [-1, 2, -1])\}.$$

Unlike PSNR, $\hat{s}$ was designed to correlate well with subjective data sets that demonstrate all the digital artifacts introduced by lossy compression. However, as shown, both PSNR and $\hat{s}$ require knowledge of the original images to determine image quality. The next section introduces a novel approach that does not require knowledge of the original images but like the $\hat{s}$ metric takes advantage of the spatial and temporal information content of the images to determine variations in quality.

4. OBJECTIVE QUALITY BASED SOLELY ON OUTPUT IMAGERY

The objective metrics introduced for transmission subsystems do not consider the other parts of the ATR system. Since the world is so dynamic, it is impossible to acquire complete end-to-end knowledge about all the analog and digital artifacts that could cause degradation to the imagery. As discussed earlier, beside the transmission system, quality is impacted by the atmosphere and obstructions between the object and the sensor, followed by the optical system of the sensor, the sensor electronics, digitization, compression, software, the receiving system, and the display system. For an image sequence, results suggest that a more general approach to quality determination is analysis of spatial and temporal vectors where the original input sequence is not explicitly given.

4.1. Approach for Obtaining Spatio-Temporal Information Content

Image sequences have varying amounts of spatio-temporal content[8] which impact the quality assessment. The spatio-temporal content of an image sequence can be plotted on a two dimensional graph of spatial content versus temporal content where values are obtained using techniques documented in an ANSI standard.[7] To obtain the spatial content, the first step is to apply the Sobel operation to each image frame (Figure 4) by centering the 3x3 Sobel mask over each pixel $Y(i, j, t_n)$, then by multiplying the mask coefficients by the neighboring pixels and finally by summing the resulting nine values together. The Sobel operation is expressed mathematically for the horizontal mask as

-1	-2	-1
0	0	0
1	2	1

-1	0	1
-2	0	2
-1	0	1

Figure 4. The 3x3 horizontal (left) and vertical (right) Sobel masks.

$$\begin{aligned} SI_h(i, j, t_n) = {} & Y(i+1, j-1, t_n) - Y(i-1, j-1, t_n) + 2Y(i+1, j, t_n) - 2Y(i-1, j, t_n) \\ & + Y(i+1, j+1, t_n) - Y(i-1, j+1, t_n) \end{aligned}$$

and for the vertical mask as

$$\begin{aligned} SI_v(i, j, t_n) = {} & Y(i-1, j+1, t_n) - Y(i-1, j-1, t_n) + 2Y(i, j+1, t_n) - 2Y(i, j-1, t_n) \\ & + Y(i+1, j+1, t_n) - Y(i+1, j-1, t_n) \end{aligned}$$

with the magnitude of the spatial information defined as

$$SI_r(i, j, t_n) = \sqrt{SI_h^2(i, j, t_n) + SI_v^2(i, j, t_n)}.$$

Next, for P total pixels in an image, the standard deviation of each Sobel filtered image is calculated as

$$SI_{stdev}(t_n) = \sqrt{\left[\frac{1}{P}\sum_i\sum_j SI_r^2(i,j,t_n)\right] - SI_{mean}^2(t_n)},$$

where

$$SI_{mean}(t_n) = \frac{1}{P}\sum_i\sum_j SI_r(i,j,t_n).$$

This results in a time series of standard deviations which can be plotted on the spatial axis of the spatio-temporal plot. To obtain the temporal information, a frame difference is computed for frame Y at time t_n defined as

$$TI(i,j,t_n) = \triangle Y(t_n) = Y(t_n) - Y(t_{n-1}).$$

The temporal information feature, $TI_{stdev}[t_n]$, is then obtained using the standard deviation of each $\triangle Y(t_n)$ calculated as

$$TI_{stdev}[t_n] = \sqrt{\left[\frac{1}{P}\sum_i\sum_j TI^2(i,j,t_n)\right] - TI_{mean}^2(t_n)},$$

where

$$TI_{mean}(t_n) = \frac{1}{P}\sum_i\sum_j TI(i,j,t_n).$$

The temporal information feature is plotted on the temporal axis of the spatio-temporal plot. The values for this feature increase with increased motion, panning, zooming and scene cuts. Figure 5 shows a spatio-temporal plot based on the first few frames of the Miss America and Table Tennis standard image sequences. The Miss America sequence has lower spatial and temporal values than the table tennis sequence. Evidence of why these values differ is shown in Figure 6 which shows more spatial and temporal activity for the table tennis sequence. The table tennis sequence has more spatial content due to textures in the scene, and it has more temporal content due to the ping pong motion and slow pan.

4.2. Spatio-Temporal Information Analysis

Figure 7 shows the spatio-temporal plot for an output sequence of 100 infrared image frames. The sequence was acquired from an infrared system mounted on an airborne platform. Evident in the result is that some frames jump out of the normal spatio-temporal area. A reformulation of the spatio-temporal information into a 1-D metric,[9] yields figure 8 which relates which frames have widely varying spatial-temporal content. The quality appears lower for those image frames that jump out of the normal spatio-temporal information content. This contention is supported by an analysis of the affect of various quality degradations on the spatio-temporal content.[9]

Further, based on a particular imaging scenario, an expected spatio-temporal information content can be predicted. In this case, not only can this technique identify individual low quality frames but it also can detect sequences of low quality frames based on a distance measure from the expected value.

Through reformulation, this metric can be the basis for a confidence factor.

Figure 9 shows a few of the infrared images which the spatio-temporal analysis suggests are lower quality. For comparison purposes, the frames before and after the degraded frames are shown. The spatio-temporal analysis obviously chose degraded images using only the output image sequence.

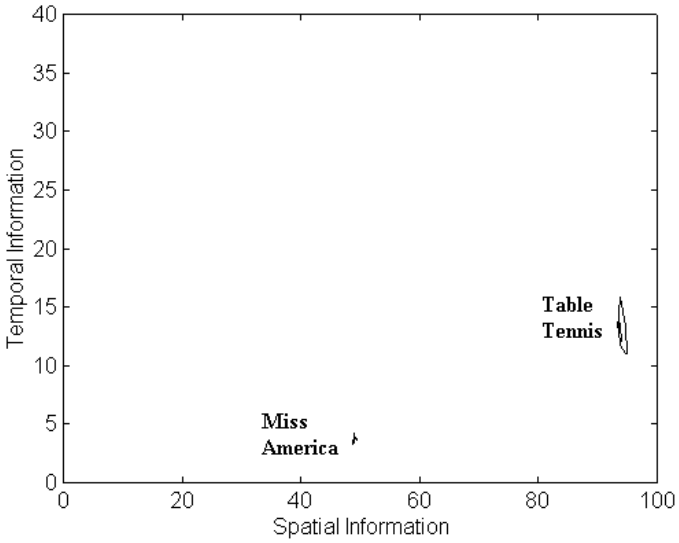

Figure 5. The Spatio-Temporal trajectories for the first ten frames of the Miss America and Table Tennis standard sequences.

Figure 6. The left two frames show original images. The middle frames show spatial content through Sobel images. The rightmost frames show temporal content via frame differencing.

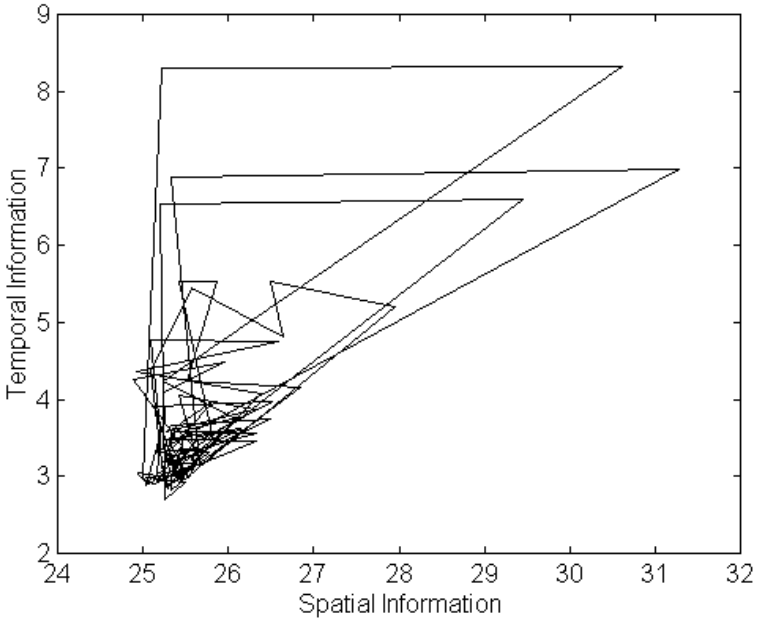

Figure 7. A spatio-temporal plot of a 100 frame infrared sequence

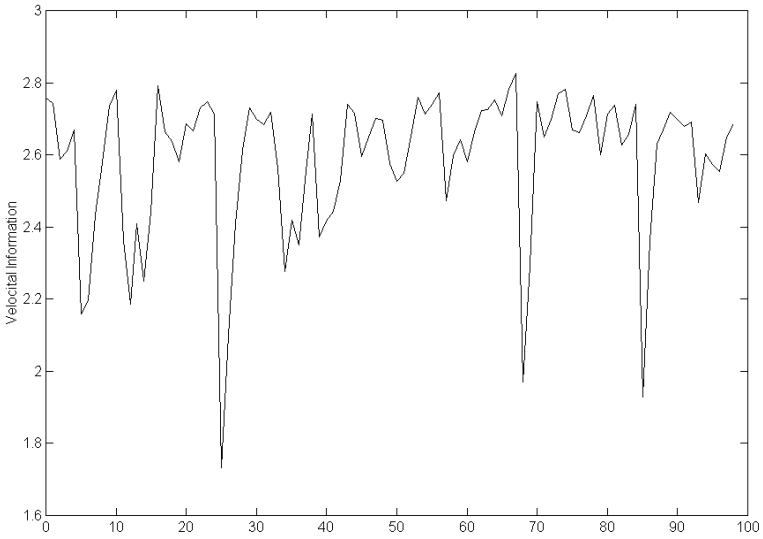

Figure 8. A reformulation of the spatio-temporal plot yields a relative quality metric per frame

5. CONCLUSION

There are numerous causes for poor image quality. Traditional approaches for measuring quality such as PSNR require knowledge of an input and output image to determine subsystem quality. In reality, since partial occlusions, atmospherics, optics and circuitry must be used to acquire an input image, there is no ideal high quality input image available. A new approach to determining quality of images in an output image sequence is introduced using no knowledge of the input image sequence. The approach is to obtain a quality metric based on an analysis of the spatio-temporal content variation in the output images. The approach was successfully demonstrated on a few infrared image sequences. The approach suggests an application to ATR where the quality metrics can be applied as a confidence factor to the ATR algorithm. The confidence factor can be obtained based on the relative quality obtained from dynamic spatio-temporal analysis. This approach would be especially useful where there is no transmission system but where the ATR system is part of the sensor, on-board a mobile platform.

The confidence factor allows a running observation of quality's impact on the recognition process. By measuring quality on the sequence, the actual probability of recognition can be raised by ignoring the lowest quality image frames or by selectively preprocessing image frames prior to applying ATR algorithms. The confidence factor can even suggest when it is necessary for humans to intervene in the recognition process.

REFERENCES

1. G. J. Power, "Motion assisted automatic target recognition," *Third ATR Systems and Technology Conference* **1**, pp. 199–204, 1993.
2. E. R. I. of Michigan, *Infrared Imaging Systems Analysis*, DCS Corporation, 1988.
3. T. Y. Young and K.-S. Fu, *Handbook of Pattern Recognition and Image Processing*, Academic Press, Orlando, Florida, 1986.
4. S. Wolf, "Features for automated quality assessment of digitally transmitted video," Tech. Rep. 90-264, US Department of Commerce, National Telecommunications and Information Administration, June 1990.
5. R. Gonzalez and P. Wintz, *Digital Image Processing*, Addison-Wesley Publishing Co., Reading, Massachusetts, 2 ed., 1987.
6. S. Daly, "Visible differences predictor: An algorithm for the assessment of image fidelity," in *Human Vision, Visual Processing and Digital Display*, *Proc. SPIE* **1666**, pp. 2–14, 1992.
7. "Digital transport of one-way video signals - parameters for objective performance assessment." ANSI T1.801.03-1996, February 1996.
8. A. A. Webster *et al.*, "An objective video quality assessment system based on human perception," in *Human Vision, Visual Processing and Digital Display*, vol. SPIE-1913, pp. 15–26, 1993.
9. G. J. Power, "Charting image artifacts in digital image sequences using velocital information content." Submitted to Proc. SPIE, July 1998.

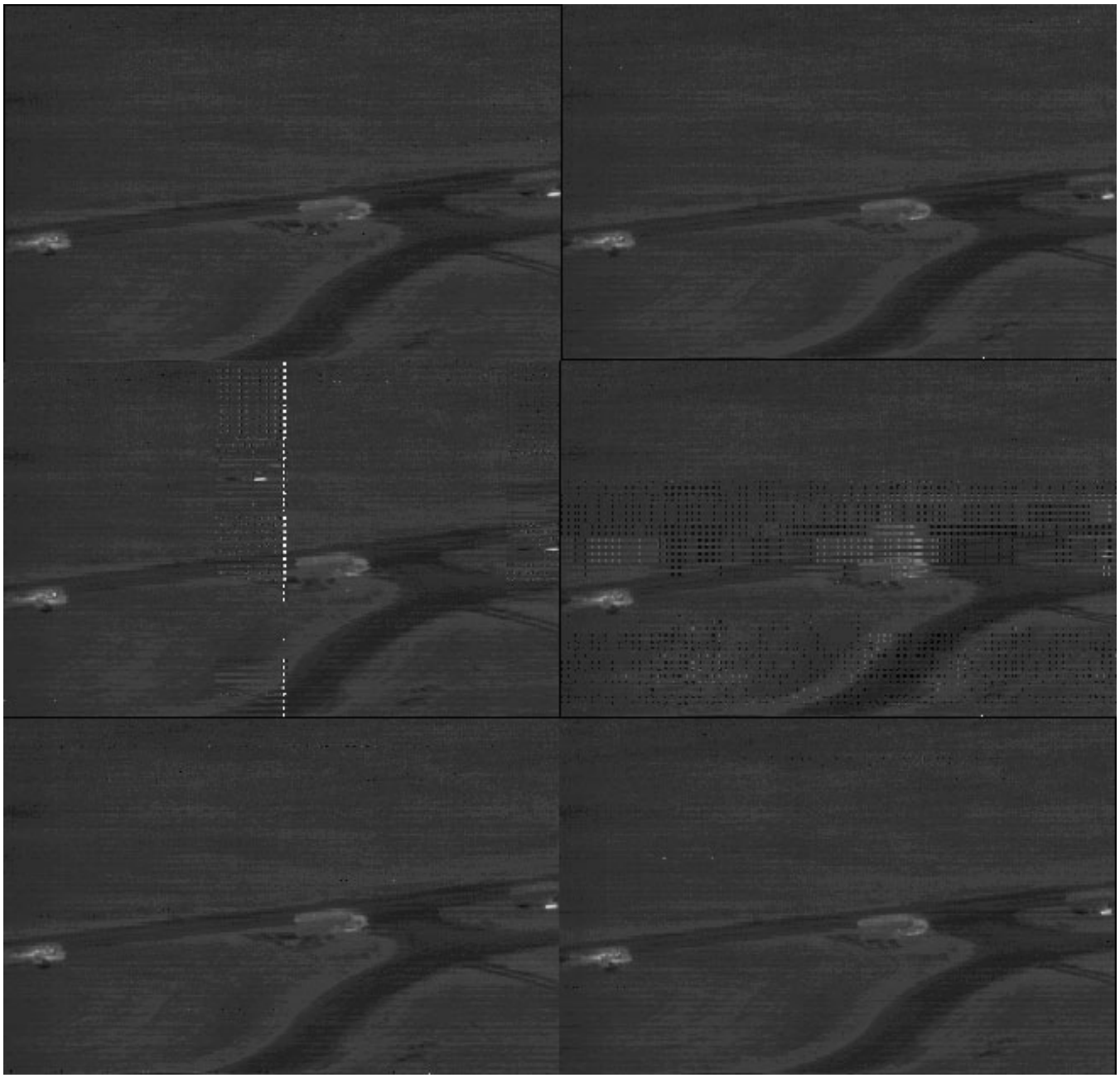

Figure 9. The middle frames were selected as lower quality via spatio-temporal analysis